On The Use of Linear Regression for the Assessment of Stability in Noise Monitoring Networks: A Practical Example

Kostas Sotirakopoulos, Richard Barham, Ben Piper, Richard Jackett
Division of Acoustics and Ionizing Radiation, National Physical Laboratory, U.K.

Summary

A practical example of a statistical method is presented for stability assessment of noise monitoring networks based on the comparison between statistical levels measured by multiple measurement stations in a network. The method makes use of an adaptation of the Chow Test and relies on the comparison of sum of squared residuals from linear regressions applied on datasets collected at different locations. It is demonstrated through application of the method on data collected near London Southend Airport that the technique does not require high temporal correlations between the tested data as long as they come from reasonably similar environments and it is shown that the same test is capable of detecting drifts form calibration when appropriate sensitivity level and time intervals are selected.

1. Introduction

The introduction of cheap components that are easy to produce in large quantities has revolutionized the modern perception of monitoring systems. Sensors like MEMS (Micro-Electro-Mechanical-Systems) microphones are being transformed from low quality commercial products to class measurement-grade equipment and their integration into measuring devices has made possible the development of noise monitoring networks at costs that seemed impossible just a few years ago [1] [2] [3] [4] [5].

However the overall costs of implementing a properly functioning sensor grid includes installation, maintenance and quality control expenses which can easily prove much higher than the manufacturing cost of the network itself. Unfortunately this implication can be a preventive factor for the implementation of such applications which in the era of information and communication become more of a necessity rather than a luxury. Hence the requirement for automated mechanisms which can guarantee good operation and longer life expectancy for measurement networks is becoming more and more pronounced. Over the last years, significant effort has been made towards addressing issues related to quality control and good operation of measuring networks. Especially in the field of anomaly detection and self-calibration various techniques have been implemented ranging from statistical and machine learning methods [6] [7] to rule based and combinatory solutions [8].

One of the most common problems met in measurement networks, which is scarcely a big issue in controlled laboratory environments, is measurement inaccuracy due to drifts from calibration. In most cases these anomalies are the result of physical fatigue of the equipment especially when low cost components exposed to public and varying weather conditions are considered. Thus the development of robust methods for the detection of such faults in networks is necessary.

In this paper an example of a statistical technique based on linear regression is demonstrated. The method makes use of an adaptation of the Chow Test which over the years has found application on various research areas. The theory and possible limitations of the test are presented and discussed.

2. Theoretical Considerations

The proposed method is based on the assumption that measurements from multiple, calibrated and normally operating nodes should follow linear patterns over time. The slopes of such patterns are depended on the length and the season of the examined time interval due to variance and seasonal characteristics found in the data. For such systems, linear regression applied over adequately long time periods should result to straight almost horizontal lines around a level that characterizes the local soundscape and does not change with time, unless of course the environment alters significantly. Assuming that the probability of synchronized and uniform drifts from calibration for multiple measurement stations is minimal, significant deviations from normal operation can be detected by comparison of the regression lines between multiple nodes.

Such comparisons can be performed by application of a variation of the Chow Test on data collected over same time periods. The introduction of the Chow Test was made by Gregory Chow in 1960 [9] and ever since then has found application on a variety of research fields; from Economics and Econometrics for the identification of structural breaks in time series, to experimental designs for regression-discontinuity analysis on psychology studies [10] [11] [12] [13]. While this test was originally developed and applied on samples coming from the same population but different time periods, it can be generalized for samples coming from different populations like data from multiple measurement positions.

The aim of the test is the statistical comparison between sets of coefficients coming from two different regressions applied on data with n and m number of observations. The initial step is the null hypothesis that there is no statistically significant difference between the two coefficient sets. Regression is applied on each observation set and the sum of squared residuals for each individual line fitting is computed. Then the two data sets are combined into a $\mathrm{n}+\mathrm{m}$ long sample and linear regression is applied again. According to Chow, the ratio between the difference of the combined sample's sum of squared residuals and the sum of the individual sum of squared residuals over the latter sum, adjusted for the appropriate degrees of freedom is distributed as an F ratio under the null hypothesis. When both samples' sizes are greater than the number of estimated coefficients this F-ratio is equal to:

$$F = \frac{(SSR_{(n+m)} - (SSR_n + SSR_m))/p}{(SSR_n + SSR_m)/(n+m-p)} \quad (1)$$

where:

- $SSR_{(n+m)}$ is the sum of squared residuals for the combined data set
- SSR_n is the sum of squared residuals for the n long set of observations
- SSR_m is the sum of squared residuals for the m long sample
- p is the number of estimated coefficients (in the case of linear regression p is equal to two since the estimated parameters are the slope and offset of the regression line)

The critical value against which the computed F ratio should be compared in order to decide on the rejection, or not, of the null hypothesis can be found in the F cumulative probability matrix for a chosen level of significance a and $(p, n+m-p)$ degrees of freedom.

For the application of the Chow test on noise data coming from different measurement locations, appropriate scaling must be applied in order to minimize discontinuities in the combined observations. That is because when comparing regression lines fitted on samples collected at different areas for the identification of anomalies, it is slope deviations that are of interest rather than the expected offset differences.

3. Examples on experimental data

In order to demonstrate the capabilities of the presented method the test was applied on experimental data collected from a trial noise measurement network, built at the National Physical Laboratory, consisting of four measurement stations which were placed around London Southend Airport. The distances between measurement nodes are shown in Table 1 while Figure 2 and Figure 1 present the units and the map locations. More information about this project can be found here [14].

Table 1: Approximate distance in meters between measurement nodes placed at London Southend Airport.

Units	*Distance (m)*
1008 - 1010	*70*
1008 – 1009	*1600*
1008 – 1012	*492*
1009 – 1010	*1500*
1009 – 1012	*2000*
1010 -1012	*560*

Figure 2: Five measurement units undergoing final testing at NPL before deployment.

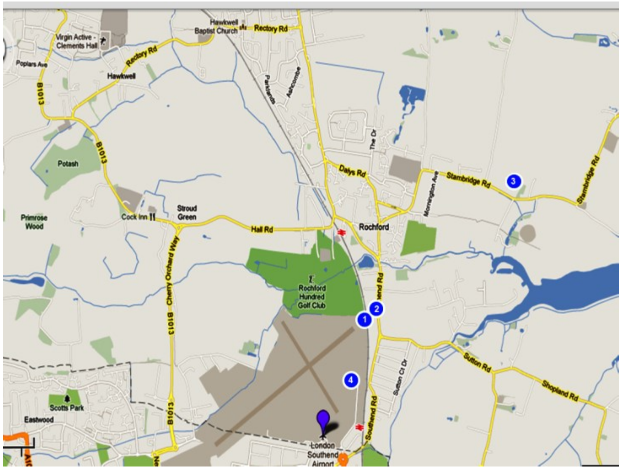

Figure 1: Location of measurement units at London Southend Airport. Units 1008, 1010, 1009 and 1012 were placed at locations 1 to 4 respectively.

The network was deployed from 1 February 2013 until 20 June 2013. During the deployment period several issues occurred ranging from voltage drops to physical weariness of the equipment, which made the collected data an interesting case study for the development and demonstration of this network stability assessment technique.

The units were set up to measure various routine parameters including broadband A-weighted L_{eq} and statistical levels with a 1 minute long measurement cycle. The readings were then saved on a server data base. Analysis of the measured parameters showed that L_{50} levels presented much less variance compared to L_{eq}. In order to further reduce the variance in the data and achieve a better fit for the linear regression and hence higher accuracy for the test, the one minute L_{50} measurements were used to compute broadband night time L_{50} levels.

Before demonstrating any comparisons it is considered useful to examine how the time series of the deployed monitoring stations look like in order to try and identify interesting features. Figure 3 presents the computed night time L_{50} levels for the whole duration of the deployment.

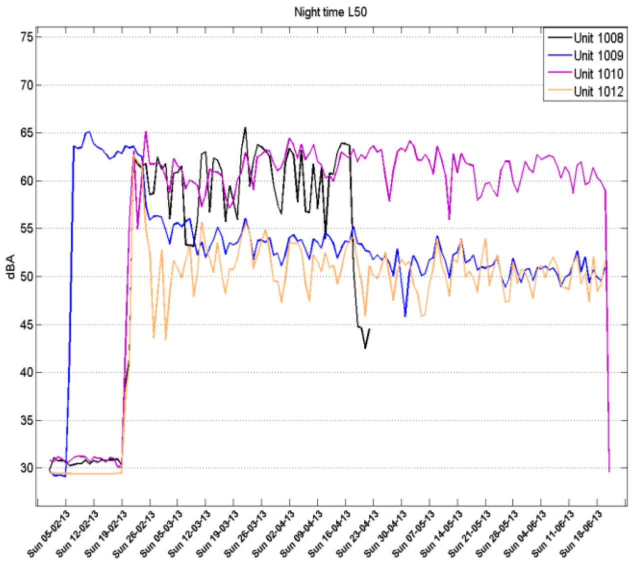

Figure 3: Night time L_{50} measurements for the whole duration of the deployment.

As seen all measurement nodes, apart from Unit 1010, presented some kind of fault during the deployment period. A gradual level decrease is observed for Unit 1009 which seems to be much more rapid for Unit 1012. This probably occurred due to faults in the power supply system of these two units which lead them to drift from calibration and finally stabilize at about 10dB above the noise floor. Moreover an abrupt noise level drop in the time series of Unit 1008 is seen. That is due to insufficient anchoring for the soft ground conditions which caused the unit to be blown over and eventually power down after a few days.

Detection of drifts from calibration

When no significant difference exists between the acoustic environment at two locations it is expected that measured noise levels will present similar behaviour with time. Obviously the closer the measurement positions are the more likely it is to observe high correlations between signals and hence the easier the identification of faults should be. However, the presented technique does not rely on strong temporal correlations in order to provide accurate detections since it examines the relation between sum of squared residuals of the applied regressions rather than point to point relations in the signals. To demonstrate this, L50 levels collected by Units 1008, 1010 and 1009 between 21 February and 09 April 2013 were tested. Figure 4.2-4 demonstrate the compared regression lines fitted on the individual and combined data sets.

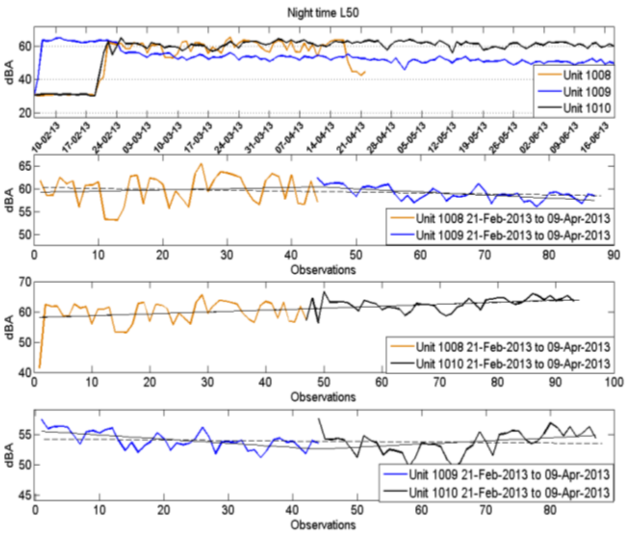

Figure 4.1: Night time L_{50} levels collected by units 1008, 1009 and 1010.
Figure 4.2-4: Comparison between Individual and combine data for units 1008-1009, 1008-1010 and 1009-1010 respectively from 21 February 2013 until 09 April 2013.

The critical value above which the null hypothesis must be rejected was computed to be equal to $F_c(2,90) = 1.63$, at a significance level $a = 0.2$, for all three tests while the F scores for the comparison between units 1008-1009, 1008-1010 and 1009-1010 were 2.71, 0.01 and 11.15 respectively. This suggests that when unit 1009 was included in the comparisons the test gave statistically significant indications for rejection of the null hypothesis that there is good agreement between regressions and thus no anomalies exist in any of the nodes. The double slopes in Figure 4.2 and Figure 4.4 confirm the results of the test. Nonetheless, the null hypothesis was far from being rejected in the test between units 1008 and 1010 which did not present any faults during the examined period.

The above test managed to produce the expected results, yet, for a number of reasons one might want to adjust the sensitivity of the comparison in order to tune the test to the acoustic characteristics of the environment. This can be implemented by selecting a different level of significance. However, something like that should be done with caution as it might eventually lead to systematic Type 1 or Type 2 errors. In the previous comparison for example the cumulative probability value of $F(2,90)$ for $a = 0.05$ is approximately 3.1 a rather lenient threshold for the specific soundscape which would introduce a Type 2 error in the comparison between nodes 1008 and 1009.

Limitations due to small network size

Particularly interesting results are obtained when the initial and steepest part of the drift for units 1009 and 1012 is excluded from the test. In order to demonstrate the effect of that, the test was applied amongst all measurement nodes from 7 March to 16 April. The critical value was found equal to $F_c(2,76) = 1.64$ at $a = 0.2$ and the results of the tests are summarized in Table 2.

Table 2: Summary of test results for tests applied between all measurement nodes over the time period 7 March to 16 April 2013.

1010 – 1012	*4.75*
1009 – 1012	*0.0004*
1009 – 1010	*9.74*
1008 – 1012	*2.26*
1008 – 1010	*0.003*
1008 – 1009	*2.66*

Table 2 shows that the only non-rejections of the null hypothesis occur when the test is applied between the normally operating nodes and those that demonstrated drifts. In order to understand why this happens one should examine how the faulty sensors operated during the examined time interval. A look in Figure 3 makes clear that over that period the sensors' readings seem to agree well while keep drifting from calibration at a remarkably similar slope. This could be due to pure luck or it could be attributed to some voltage stabilizing system embedded in the power supply unit to avoid shut downs due to voltage drops. In any case, if the initial part of the slopes is excluded, then the assumption made at the start, that it is very unlikely for multiple units to present faulty behaviour in an almost identical way, is violated and in small networks, like the one examined here, with very few nodes this could lead to inability to draw any reasonable conclusions on which sensors continue operating normally and which not.

4. Conclusions

The practical examples presented in this paper demonstrate how the Chow Test can be adapted for the evaluation of noise monitoring networks' stability through comparison between noise levels recorded at different locations. It was shown that close proximity of measurement nodes and hence high network density is not necessary for this technique to perform well so long as the soundscape characteristics do not vary excessively between measurement locations, a condition which holds for the examined area around London Southend Airport. When the test is applied over appropriate time periods it can prove a robust method for the detection of drifts from calibration. Moreover tuning of the test to fit the acoustic characteristics of the soundscape is possible by adjustment of the critical value through the level of significance. As discussed this is a procedure that requires some experience since it can lead to systematic errors in the results. Finally examination of the robustness and accuracy of the demonstrated method at urban environments, where soundscape variations between locations can be much greater, is considered to be an interesting next step.

Acknowledgement

This project has been funded by National Measurement Office of the United Kingdom Department for Business, Innovation and Skills.

References

[1] R. Barham, M. Goldsmith, M. Chan, D. Simons, L. Trowsdale, R. Tyler: Development and performance of a multi-point distributed environmental noise measurement system using MEMS microphones, Proc. Euronoise 2009, Edinbrough, 2009.

[2] G. Memoli, D. Dawson, D. Simmons, R. Barham, M. Hamilton, R.M. Grounds, B. Philips: Towards the acoustical characterization of an Intensive Care Unit, Applied Acoustics 79 (2014), 124-130.

[3] Richard Barham, M. Chan, M. Cand: Practical experience in noise mapping with MEMS microphone based distributed noise measurement system, Proc. InterNoise 2010, Lisbon, 2010.

[4] Barham, R: New Paradigm for Environmental Noise Measurement, Proceedings of the Institute of Acoustics, Birmingham, 2014.

[5] T. Van Renterghem, P. Thomas, F. Dominguez, S. Dauwe, A. Touhafi, B. Dhoedt, D. Botteldooren: On the ability of consumer electronics microphones for environmental noise, Journal of Environmental Monitoring, vol. 13, no. 3 (2010), 544-552.

[6] V. Chandola, A. Banerjee, V. Kumar: Anomaly detection: A survey, ACM Computing Surveys, vol. 41, no. 15 (2009), 1-58.

[7] Y. Zhang, N. Merantia, P. Havinga: Outlier detection techniques for wireless sensor networks: A survey, Communications Surveys and tutorials, IEEE, vol. 12, no. 2 (2010), 159-170.

[8] S. Dauwe, D. Oldoni,B. De Baets,T. Van Renterghem, D. Botteldooren, B. Dhoedt: Multi-Criteria Anomaly Detection in Urban Noise Sensor Networks, Environmental Science : Processes & Impacts, vol. 16, no. 10 (2014), 2249-2258.

[9] G. C.Chow: Tests of equality between sets of coefficients in two linear regression, Econometrica, vol. 3, no. 26 (1960), 591-605.

[10] F.M. Fisher: Tests of equality between two sets of coefficients in two linear regression: An expository note, Econometrica, vol. 38, no. 2 (1970), 361-366,.

[11] T.D. Cook, D.T. Campbell, Quasi-experimentation: Design & analysis issues for field settings, Rand McNally & Co, Chicago, 1979.

[12] J.P. Braden, T.J. Bryant: Regression discontinuity designs: Applications for school psychology, School Psychology Review, vol. 19, no. 2 (1990), 232-239.

[13] H.B. Lee: Using the Chow Test to Analyze Regression Discontinuities, Tutorials in Quantitative Methods for Psychology, vol. 4, no. 2 (2008), 46-50.

[14] R. Jackett, R. Barham: Development of the Minim Outdoor Noise Monitoring System, Report AC9, National Physical Laboratory, London, 2014.